

Wide Area Measurements of Voice Over IP Quality Revisited

Ian Marsh¹ and Fengyi Li²

¹Swedish Institute of Computer Science, Kista S-164 29, Sweden. E-mail: ianm@sics.se.

²KTH IMIT, Kista S-164 40, Sweden. E-mail: d97-fli@nada.kth.se.

Abstract

Time, day, location and instantaneous network conditions largely dictate the quality of Voice over IP calls. In this paper we present results of VoIP quality mainly in terms of packet jitter and loss. This work revisits our objective of measuring VoIP quality in 1999 enabling us to make comparisons with the results we obtained three years ago. It also allows us to improve the measurement infra-structure we used previously through adding new functionality plus using more test sites. Furthermore we present findings on the influence of parameters such as packet size, asymmetry, silence suppression and location as well as much more comprehensive loss and jitter results. Our results conclude that the quality of Voice over IP sessions have improved over the last three years. In spite of this fact however, achieving acceptable *global* Voice over IP quality still remains a challenge.

1 Introduction

No standardised scheme exists for monitoring VoIP quality. It is well known that VoIP users of the Internet are susceptible and sensitive to quality changes, causing them to abandon calls when the quality deteriorates below an acceptable level [1]. The quality of VoIP sessions can be often quantified by packet delay, loss and jitter.

A wide area monitoring scheme, managed over multiple domains, would be beneficial to allow “administrators” to adjust resources for users of an Internet telephony service. Therefore the objective of this work is to describe a VoIP quality monitoring scheme, summarise results we have obtained, plus motivate how to choose certain sites, packet sizes, sampling frequencies and the use of silence suppression. We performed the measurements continuously over a number of weeks, providing us with more than 10000 simulated phone calls. It should be stressed that our work concentrates on the data transport of VoIP sessions, it does not include any call signalling in the measurements.

Using off-line techniques to probe and monitor the Internet for quality changes is non-trivial. By off-line we mean periodically sending pre-recorded conversations and measuring their received quality. It is non-trivial as exact call durations are not well known when there is no pay-

ment structure. Also no restrictions are made on the voice coding thus resulting in non-standard packet sizes being used, particularly by commercial vendors. Our contribution in this area is to motivate where choices of measurement parameters such as packet size, probe duration, network asymmetry can affect the results, in addition to collating a large repository traces from widespread sites. We also state the quality of connections in 2002 and compare them with those we gathered in late 1999.

This paper starts with a brief description of the current measurement infra-structure in Section 2. We show the locations of the hosts used now and those used three years ago. Also included are details of the sample file we used to probe connections between sites and a summary of data we gathered as a result of this probing. Our measurement efforts form the main body of this paper and are presented in Section 3, with particular attention to loss and jitter measurements. The related work section compares results obtained by other researchers in similar efforts and ours. Finally, we round off the paper with conclusions and a short discussion in Section 5.

2 Simulating IP Telephony Calls

Our method to measure call quality is to send pre-recorded calls between participating sites. Through modification of our own VoIP tool (called Sicsophone) we probe the intervening connections for variations in VoIP quality. We have carefully chosen nine sites, with large variations in both Internet hops and geographic distance to obtain a wide cross-section of distributed points. The location of the sites is shown in Figure 1.

One of the major changes we made since the last measurements [2] was to fully inter-connect the hosts we had available to us. Nine hosts connected to each other allows us eight connections to the other hosts giving a possible 72 interconnections. This partly accounts for the much larger amount of traces we were able to obtain. In almost all the connections tested we initiated a bidirectional call, simultaneously joining two parties. This is also an improvement over measurements made in 1999 where most sessions were only unidirectional. This scheme is also responsible for the large number of sessions we have.

We found that ten (of the 72) routes gave an unequal number of hops between the same hosts in opposite direc-

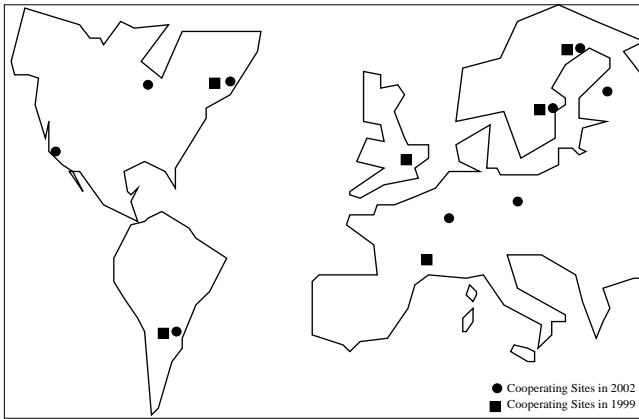


Figure 1: Locations of Test Sites

tions. Route asymmetry can be troublesome for IP telephony as one party might be experiencing different quality than the other, possibly causing one party to terminate the connection. We could not test the full mesh for asymmetry due to restrictions on the use of traceroute and firewalls. The frequency of calls was chosen to be one per hour between each cooperating sites, due to load considerations on remote machines. For most of our measurements we found this sufficient, especially when repeated over several weeks. In some cases, we performed simultaneous tests to assess whether schemes such as silence suppression or packet size had an impact. In these cases four or more connections were conducted in parallel. Table 1 summarises the call characteristics in the top half and the data we obtained from it in the bottom half¹. We purposely did not change the trace file used in 1999 to keep comparisons simple and point out that other researchers have investigated the affect of call duration on quality (Section 4).

Table 1: Used and Obtained Data

| Used Data | Details |
|-----------------------|--------------|
| Call duration | 70 secs |
| Silence Sup. | 2043 packets |
| No Silence Sup. | 3653 packets |
| Coding | 8 bit PCM |
| File Size | 584480 bytes |
| Obtained Data | |
| Total Trace Files | 10288 |
| Bidirectional Tests | 3742 |
| Differing Packet Size | 2491 |
| Suppression Tests | 1251 |
| Measurement Duration | 8 weeks |

¹ Available from <http://www.sics.se/~ianm/COST263/cost263.html>.

3 Results

3.1 1999 and 2002 Differences

This study is a follow up of work we started in 1999 [2]. We concluded then most calls were completed with a loss percentage less than 5% and a network delay below 150 ms. These results still hold in Spring 2002. For the hosts outside South America, all but one exhibited a delay less than 100ms. The average one-way delay between Argentina and the other sites is 133ms. Broadly speaking we have seen an improvement in the quality of VoIP sessions over the *same* connections. Table 2 shows the delay, jitter and loss for all the traces we obtained and their standard deviations in brackets.

Table 2: 1999 and 2002 VoIP Quality Differences

| Quality | 1999 | 2002 | % Diff. (+/-) |
|---------|--------|-------------------------|---------------|
| Jitter | 45.1ms | 22.6ms (± 13.7) | -50.0% |
| Loss | 1.2% | 0.5% (± 2.6) | -58.3% |
| Delay | 115 ms | 84.95ms (± 44.85) | -26.1% |

The remainder of the results presented refer to our measurements taken in 2002.

3.2 Loss

Generally the measurements produced very good quality. 75% of the conversations showed zero percent loss and over 96% with a loss rate less than 10%. We quote 10% as speech coding schemes with tolerance to lost data, forward error correction mechanisms and human tolerance. Figure 2 shows a typical loss pattern over one day for two different connections, the x-axis starts from 0:00 up to 24:00. Note we use our local time (Swedish) to be able to plot the two on the same figure. In relative terms the loss differences is large but both are below 0.5% packet loss. Connections to and from Europe/USA and South America were slightly better than we reported previously with an average loss rate of just over 3%.

Figure 3 shows the consecutive losses for all of the traces. The y-axis is a \log_{10} scale. The majority of losses are one to four consecutive losses. We also observed some very long loss sequences almost inevitably caused by temporary outages at either the sender or the receiver. From examination of the trace files they were nearly always at the end of the session resulting in no more packets being sent or received. We therefore cannot attribute these losses to dropped packets in inter-mediatory routers. The number of files exhibiting these large losses is small compared to the 10000 files we collected, and 40-50 packets is still only 1.5% of the size of the sample call.

3.3 Jitter

Most traces exhibited low jitter values. The average across all files sites was 22.6 ms ($\pm 13.7ms$). Figure 4 shows a

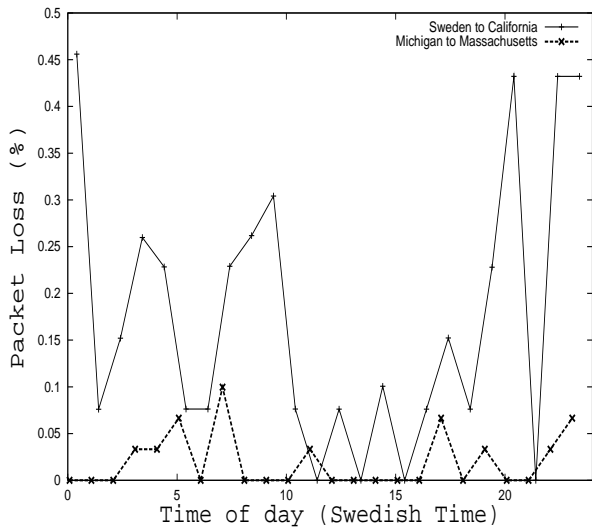


Figure 2: 24 hour Loss in 2002

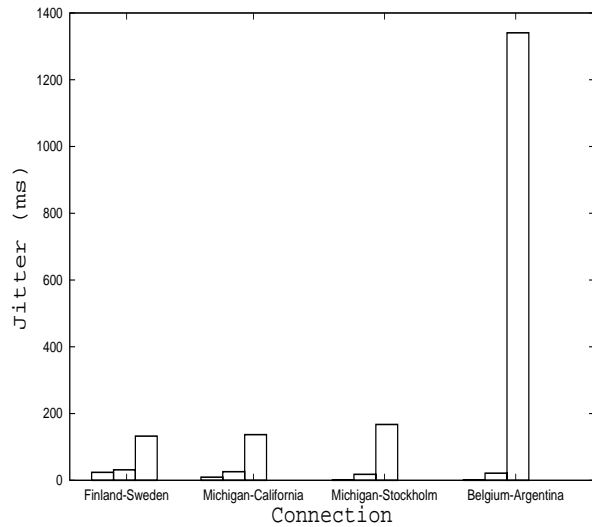


Figure 4: Jitter for Selected Sites

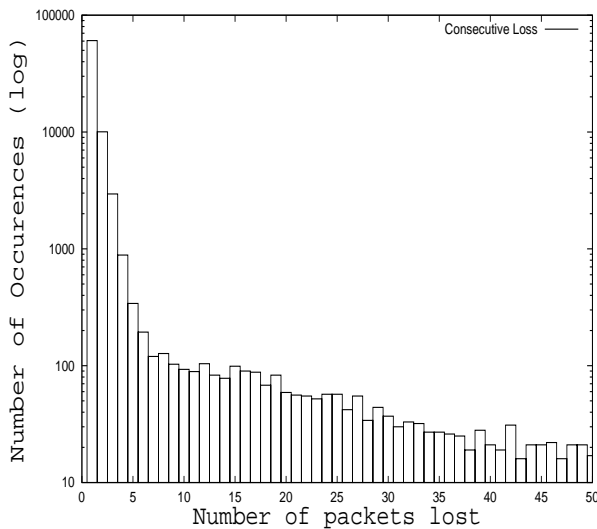


Figure 3: Consecutive Loss

jitter plot for four of the connections. The plot shows three columns, the minimum, average and maximum jitter for a connection. The first labelled site is the sender and the second the receiver (i.e. from Finland to Sweden). The jitter values are an average of all the sessions we recorded. None of the connections were tested less than 400 times, approximately 700 was typical. In the rightmost two plots the minimum jitter was so small that it is barely noticeable. The difference *to* the South American site and the others however is clearly evident. This message is repeated in the next experiment we conducted.

Figure 5 shows the sites grouped by location. In this plot we show the maximum jitter experienced by each chosen connection. The key is ordered about the plot, Massachusetts to Michigan is the first column and so on for all 16 data sets. Most columns contain between 100 and 300 data points. The dispersion of the data is the facet we want

to highlight. Quite clearly the message of poor quality *to* South America is repeated in this plot, interesting to note is the maximum jitter *from* Buenos Aires is clearly much less than *to* it. Unfortunately the host in Argentina was one which we could not perform a traceroute from it.

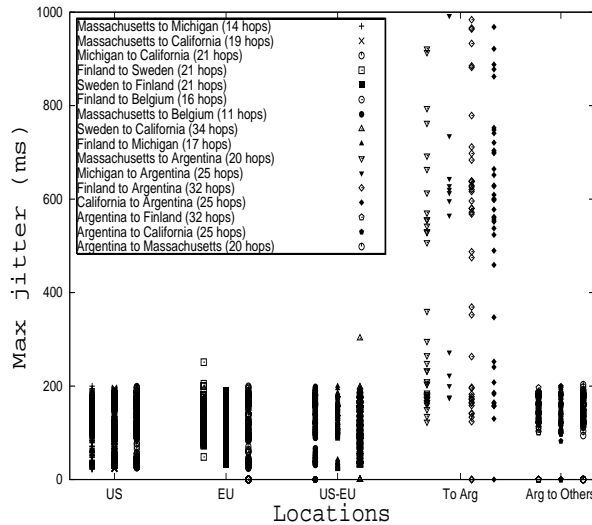


Figure 5: Maximum Jitter per Location

3.4 Network Asymmetry

By network asymmetry we mean the connections from one site to another is not the same in the opposite direction. This can result in differing delays, loss rates and jitter for data on the Internet [3]. In our previous work we concluded that network asymmetry could play a significant role in the perceived quality of one of the callers. We found that on average there was an asymmetry. However it is not so serious in terms of VoIP quality as reported for

the traces we compared. The loss values were up to 0.5-1.25% different for calls taken at the same time in opposite directions. An example is shown in Figure 6.

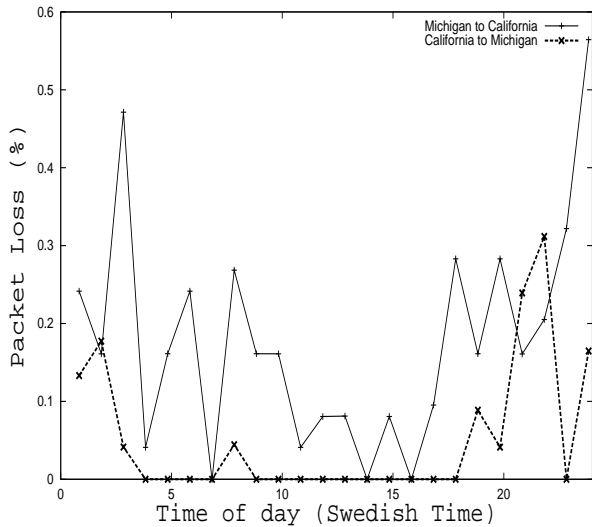


Figure 6: Route Asymmetry

3.5 Silence Suppression

Measuring loss and jitter were not the only goals of this work. In our previous measurements we did not address the effect of silence suppression on measuring call quality. Not using silence suppression (or a sample file with no silence periods) enables more sampling of the network quality, as packets are sent even when the sender is silent. The cost is however more load on the network. Table 1 shows that only 56% of the packets need to be sent using silence suppression, a significant saving on a congested network. Figures 7 and 8 show the jitter of two simultaneous calls at the receiver, the top plot not using silence suppression and the bottom one utilising it. The differences are small, slightly more jitter is experienced by the call using silence suppression, indicated by the spikes, we can attribute this to a slight timing mismatch of the two calls, as the packets are not perfectly interleaved, possibly allowing data packets to be inserted or the streams to interfere with each other. Our recommendation is not to use silence suppression where no loss is being experienced, and use silence suppression if losses are being encountered. Loss in itself was not affected when using silence suppression over not using it (packet size 160 bytes).

3.6 Packet Size

Another factor is the choice of packet size. Finding the optimal size for measurement partly depends on the applications that will use the network in an operational setting. Different codings, applications and standards can dictate the final choice of packet size. Some applications are even

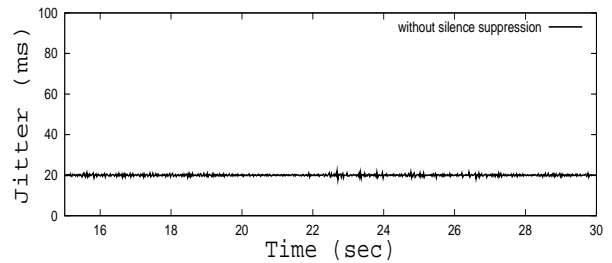


Figure 7: Without Silence Suppression

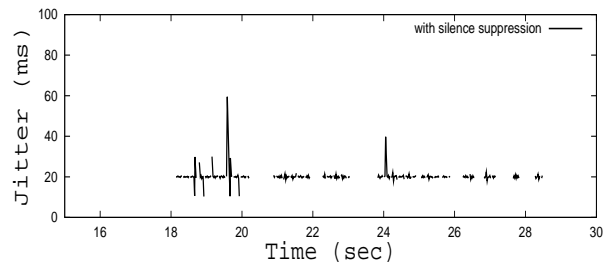


Figure 8: With Silence Suppression

capable of selecting different coding styles resulting in different packet sizes. In order to assess if a correlation exists between packet size and the loss rate we ran several parallel experiments with packet sizes ranging from 160 to 1280 bytes. Our conclusion is that no noticeable difference in loss was experienced. Figure 9 shows the loss rates between using 160 and 1280 bytes over a 24 hour period. One drawback with using large packets is less data is produced for sampling due to larger chunks, but the overall load is lower due to less space used for each header (20 IP + 8 UDP and 12 RTP = 40 bytes).

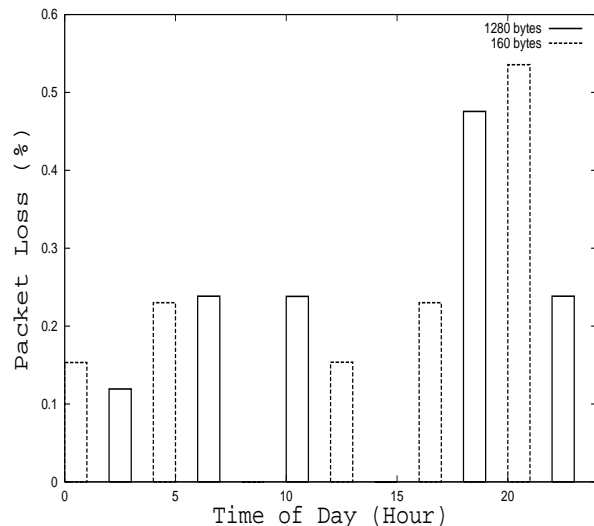


Figure 9: Effect of Packet Size

4 Related Work

Bolot and Hardman looked at consecutive loss, for a FEC scheme [4] [5]. They concluded that the number of consecutive losses is quite low, Bolot stated that between one and five at 8am and between one and ten at 4pm. Our results (shown in Figure 3) are in agreement with these findings. We also have seen traces with more losses, but we had many more traces and can attribute some losses due to reasons other than buffer overflow in routers. Maxemchuk measured both loss and delay variation for intra-state connections, within the USA and international links [6]. Their conclusion was the quality depends on the length of the connection and the time of day. We did not try different connection lengths but found variations during a 24 hour cycle (shown in Figures 2 and 6). It is worthy to point out our loss rates were considerably less than theirs (3-4%). Dong Lin had similar conclusions in [7], stating that in fact even calls within the USA could suffer large jitter delays. Her results on packet loss also agree with those in [4], which is interesting, as the measurements were taken four years later. Her thesis also contains results on trying different packet sizes and concludes that loss patterns were almost indistinguishable for packet sizes from 5 to 512 bytes. We also did not observe very differing losses for packet sizes of 160, 640 and 1280 bytes, as shown in Figure 9.

5 Conclusions and Discussion

Our second measurement attempt took in more hosts, nine rather than six and more importantly connected them in a full-mesh configuration. We now have over 15000 simulated calls (10000 were used in the preparation of this paper) compared to several hundred in 1999. Performing measurements on this scale is not an easy task. Different access mechanisms, firewalls, NATs even non super-user access complicates the work to obtaining measurements.

Voice over IP is still feasible in 2002, despite the dramatic rise of WWW traffic. The majority of calls we tested exhibited excellent quality; the vast number of calls completed with less than 5% packet loss and under 50ms jitter. Compared to 1999 we have seen an improvement in quality of telephony communication using the Internet. The delay has improved by 26% and the jitter and loss by over 50%. We can only attribute this Internet Service Providers upgrading and improving their networks. We have seen higher network asymmetry, measured in hops, between the same sites we used in 1999 and 2002, but with the exception of one host has not significantly affected our results. We still see a disparity between hosts in the EU countries and the USA and those outside these areas. Despite our careful choice of sites in 1999 and 2002 we still see a stark separation into two categories. We therefore plan to investigate this further and have obtained access to sites in Turkey and New Zealand. We have also observed factors

such as packet size and the use of silence suppression do not have a significant impact of quality of Internet connections.

References

- [1] D. De Vleeschauwer, J. Janssen, G. H. Petit, and F. Poppe. Quality bounds for packetized voice transport. *Alcatel Telecommunications Review*, pages 19–24, 1st Quarter 2000.
- [2] Olof Hagsand, Kjell Hansson, and Ian Marsh. Measuring internet telephone quality: Where are we today ? In *Proceedings of the IEEE Conference on Global Communications (GLOBECOM)*, Rio, Brazil, November 1999. IEEE.
- [3] Vern Paxson. End-to-end routing behavior in the internet. In *SIGCOMM Symposium on Communications Architectures and Protocols*, Stanford, California, August 1996.
- [4] J. Bolot, H. Crepin, and A. Garcia. Analysis of audio packet loss in the internet. In *Proc. International Workshop on Network and Operating System Support for Digital Audio and Video (NOSSDAV)*, Lecture Notes in Computer Science, pages 163–174, Durham, New Hampshire, April 1995. Springer.
- [5] Vicky Hardman, Angela Sasse, Mark Handley, and Anna Watson. Reliable audio for use over the internet. In *Proc. of INET'95*, Honolulu, Hawaii, June 1995.
- [6] N. F. Maxemchuk and S. Lo. Measurement and interpretation of voice traffic on the internet. In *Conference Record of the International Conference on Communications (ICC)*, Montreal, Canada, June 1997.
- [7] Dong Lin. Real-time voice transmissions over the internet. Master's thesis, Univ. of Illinois at Urbana-Champaign, 1999.